

# Measures of Teacher Effectiveness in Teacher Evaluation Systems

**November 2011**

**Alicia N. Garcia, J.D.**  
**Jennifer L. Stephan, Ph.D.**



**REL Midwest**

1120 East Diehl Road, Suite 200  
Naperville, IL 60563-1486  
866-730-6735  
<http://edlabs.ed.gov/RELmidwest/>

This memorandum was prepared for the Institute of Education Sciences (IES) under contract ED-06-CO-0019 by Regional Educational Laboratory Midwest, administered by Learning Point Associates, an affiliate of American Institutes for Research. The content does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government. This document is not verified as meeting IES standards and is not intended for public distribution.

1415\_11/2011

# Contents

	Page
Introduction.....	1
Overview.....	1
Guidelines in Measuring Teacher Effectiveness.....	2
Approaches to Evaluating Student Growth for Teachers in Nontested Subjects.....	3
Other Measures Used to Assess Teacher Effectiveness .....	6
Summary.....	8
Summary of State Policies on the Evaluation of Teachers in Nontested Subjects .....	9
References.....	13

## Introduction

REL Midwest prepared this brief in response to a request from the Michigan Career and Technical Education Research and Data Use Committee. The committee seeks to better understand options for evaluating career and technical education (CTE) teachers that are consistent with recent policy initiatives. The committee requested that REL Midwest provide background information on what types of measures are used in teacher evaluation systems, particularly to evaluate teachers in nontested subjects.

In the following section, this brief describes key guidelines that have emerged from district, state, and federal policy initiatives for measuring teacher effectiveness. Then it details the types of measures used in teacher evaluation systems with a particular focus on student growth measures for teachers in non-tested subjects. The information was found by searching the National Comprehensive Center for Teacher Quality website and Google Scholar using the terms “non-tested,” “nontested,” and “career and technical” in conjunction with “teacher evaluation.” The overview focuses on documents published in 2009 or later.

This brief concludes with a state-by-state scan of how states evaluate student growth for teachers in nontested subjects. The state-by-state scan includes states in the Midwest region. Tennessee and New York are also included because the Michigan Department of Education specifically requested information on these two states. The information presented in the state-by-state scan was obtained primarily through a search of state education agency websites. An Internet-based search was conducted primarily using Google, relying on search terms that included the name of a state and “teacher evaluation,” “CTE teacher evaluation,” and “teacher evaluation nontested subjects.”

## Overview

In response to recent federal (e.g., Race to the Top, the Teacher Incentive Fund) and state initiatives, developing a comprehensive teacher evaluation system has become a priority for many states. Michigan passed legislation in 2010 and 2011 to create a teacher evaluation system that links teacher effectiveness ratings to tenure, promotion, and retention decisions (Public Act No. 205, Section 1249). While research has repeatedly demonstrated the importance of teachers to students’ achievement (Darling-Hammond, 2000; Hanushek & Rivkin, 2010; Nye, Konstantopoulos, & Hedges, 2004; Rockoff, 2004), which aspects of teaching matter most and how to measure them is less understood, and states and districts vary in how they measure teacher effectiveness. One key component of many evaluation systems is a measure of student growth in performance on state standardized assessments. However, in most districts, the vast majority of teachers do not teach in a tested subject or grade (Glazerman et Al., 2011; Kane, Taylor, Tyler, & Wooten, 2011; Prince et al., 2009).<sup>1</sup> This overview describes guidelines that

---

<sup>1</sup> In a widely cited report, Prince, Schuermann, Guthrie, Witham, Milanowski, & Thorn (2006) coined the term “the other 69 percent” to refer to teachers of nontested subjects and grades. This term derives from the fact that in 2004-

have emerged from district, state, and federal policy initiatives and practice for selecting evaluation measures. It then reviews the types of measures used in evaluation systems for teachers in both tested and nontested subjects with a focus on student growth measures for teachers in nontested subjects because these present some challenge for evaluating CTE teachers.

## **Guidelines in Measuring Teacher Effectiveness**

Previous research has identified four principles consistently represented in recent teacher evaluation policy or practice. These principles may serve as guidelines as other districts and states develop teacher evaluation systems.

- (1) Multiple measures should be used to evaluate teachers (Glazerman, et al., 2011; Potemski et al., 2011; Reform Support Network, 2011; Steele, Hamilton, & Stecher, 2010).

Different measures capture different aspects of teaching. For example, process measures (e.g., classroom observations) reflect the interactions of teachers and students within a classroom, while output measures (e.g., student growth on standardized tests) reflect the impacts of teaching (Goe, Bell, & Little, 2008). Different measures can also provide information that serves different purposes. Summative measures are linked to employment decisions, while formative measures provide detailed information useful for improving teacher practice. For example, while classroom observations can be used in either a formative or summative way, measures of student growth on state standardized exams provide little information specific enough for professional development (Mather, Oliva, & Laine, 2008). In addition to providing information on various aspects of teaching and for different purposes, using multiple measures can increase the reliability of evaluations.

- (2) Student growth measures are a key component of teacher evaluation systems (Glazerman, et al., 2011; Goe & Holdheide, 2011; Potemski, Baral, & Meyer, 2011; Reform Support Network, 2011). Student growth refers to the change in a student's achievement over two or more points in time. When used in a teacher evaluation system, student growth measures assign a score to a teacher (or team of teachers) based on students' growth in achievement.<sup>2</sup> Growth in achievement, rather than achievement alone, is used to account for differences in students' prior achievement. Several different models have been developed to measure student growth using state standardized assessments—see Braun (2005) for a discussion. This brief will discuss approaches for measuring student growth

---

05, only 31 percent of teachers in Florida taught in reading or math. Kane and colleagues (2011) estimate that 75% of K-12 teachers are in a nontested grade or subject.

<sup>2</sup> Which students should count towards the score for a particular teacher is a particular concern for caseload teachers, such as special education teachers. See Holdheide, Goe, Croft, & Reschly (2010) and Goe & Holdheide (2011) for a discussion.

in nontested subjects. In Michigan, student growth will account for 25 percent of a teacher's evaluation rating in 2013-2014, 40 percent in 2014-2015, and 50 percent in 2015-2016 (Metz & Wraight, 2011).

- (3) Measures should be rigorous and comparable (Goe & Holdheide, 2011; Reform Support Network, 2011). Rigorous measures are measures that evaluate knowledge and skills appropriately (e.g., against grade-level standards), and comparable measures are measures that can be used to compare teachers across classrooms in a district or state (Goe & Holdheide, 2011).
- (4) Measures should differentiate among teachers (Glazerman et al., 2011; Reform Support Network, 2011). A prominent criticism of traditional teacher evaluations is that they are undifferentiated: the vast majority of teachers receive superior ratings (Glazerman et al., 2011; The New Teacher Project, 2010; Toch & Rothman, 2008). In Michigan, teachers will be rated as "highly effective," "effective," "minimally effective," or "ineffective," and measurements need to be sensitive enough to differentiate teachers across these categories.

These four principles have emerged as priorities across districts and states that are designing teacher evaluation systems, including Michigan, and can be considered when proposing specific measures to evaluate CTE teachers in Michigan.

For CTE teachers, however, the directive to measure student growth (guideline 2) may present some challenge. State standardized assessments developed under the No Child Left Behind Act (NCLB) can be used to measure student growth (guideline 2) in a comparable way across classrooms (guideline 3) for tested grades and subjects, but CTE teachers, and indeed the majority of teachers in many districts, do not teach in a subject measured by state assessments created under NCLB (Glazerman, et. al, 2011; Kane, Taylor, Tyler, & Wooten, 2011; Prince, et al., 2009). What approaches do districts and states use to measure student growth in nontested subjects? Beyond student growth, what types of measures have been included in teacher evaluation systems? The following section considers these two questions by reviewing the types of measures used to assess student growth and evaluate teachers more generally. Some advantages and disadvantages are discussed. This brief then concludes with a summary of state approaches to evaluating student growth for teachers in nontested subjects.

## **Approaches to Evaluating Student Growth for Teachers in Nontested Subjects**

Including a measure of student growth is one of the key guidelines in measuring teacher effectiveness (guideline 2). Three general approaches to measuring student growth in nontested subjects, described below, have emerged from district and state practice (Reform Support Network, 2011). The advantages and disadvantages of each, with special attention to guidelines 3 (measures should be rigorous and comparable) and 4 (measures should differentiate among teachers), are discussed below.

## Student Learning Objectives (SLOs)

In evaluation systems that use SLOs, a teacher works with an administrative leader to set measurable goals at the beginning of the year. At the end of the year, the administrator evaluates whether the teacher's goals have been met. While SLOs can be based on state standardized assessments, they are useful for teachers of nontested subjects because they can also be based on teacher-developed assessments (Race to the Top Technical Assistance Network [now known as the Reform Support Network], 2011; Goe & Holdheide, 2011). A music teacher, for example, could have an SLO based on how much students improve their performance of the same piece of music over time (Goe & Holdheide, 2011).

SLOs are one part of the teacher compensation system in Denver Public Schools (ProComp), where teachers develop two student growth objectives annually. CTE teachers, for example, are directed to write one objective based on student growth in the course content area and another objective that can focus on either student growth on a specific unit in the content area or on growth for a specific student subgroup (Denver Public Schools, n.d.).<sup>3</sup> Other districts using this approach are located in Charlotte-Mecklenburg, North Carolina, and Austin, Texas (Race to the Top Technical Assistance Network, 2011). Beginning with the 2011–12 school year, Rhode Island has adopted SLOs as part of its teacher evaluation model (Rhode Island Department of Education, 2011).

This approach has several advantages. It can be adapted to a variety of subjects, types of teaching assignments, and assessment structures, and it directly involves teachers in designing goals (Goe & Holdheide, 2011; Race to the Top Technical Assistance Network, 2011). On the other hand, this approach is time intensive for teachers and administrators, and SLOs can vary in their rigor (guideline 3) and the extent to which they differentiate among teachers (guideline 4) depending on how they are written by teachers and administrators. Moreover, comparability across classrooms or schools (guideline 3) may be difficult to achieve (Goe & Holdheide, 2011; Race to the Top Technical Assistance Network, 2011). Providing specific choice parameters may improve comparability (Steele et al., 2008), rigor, and how well SLOs differentiate among teachers.

## New or Existing Alternative Assessments

States or districts could create new or use existing assessments to evaluate subjects not covered by the state assessments developed under NCLB. For example, Hillsborough County, Florida, has developed end-of-course exams over the last two decades to meet district requirements for

---

<sup>3</sup> While the documentation from Denver Public Schools does not provide an example for CTE teachers, a sample objective is provided for a teacher in a different non-tested subject, fitness. A sample objective for a fitness teacher is: “80% of students who attend 85% of the time will improve at least 10% in the cardiovascular fitness as measured by one of two assessments:

- Pacer Test
- 600-Yard Run Test” (Denver Public Schools, n.d., p. 44).

testing students in all subjects and grades. In May 2007, there were more than 500 tests, many including a pretest, that cover a diverse list of subjects including art, physical education, and career and technical education. Hillsborough County made these exams available to other districts in Florida between 2007 and 2009 through a clearinghouse, but it is no longer funded (Steele et al., 2010). States in the process of identifying or developing tests for typically nontested subjects include Delaware (Race to the Top Technical Assistance Network, 2011) and Tennessee (Tennessee Department of Education, 2011).

This approach can be adapted to a variety of subjects and can yield rigorous and comparable (guideline 3) results (Race to the Top Technical Assistance Network, 2011) that differentiate among teachers (guideline 4). For CTE, assessments already exist in some areas because Perkins IV requires states to report a measure of technical skills attainment. However, there are also important considerations, and potential disadvantages, to this approach. In some cases, commercial assessments can be purchased to assess subjects not tested under NCLB. Commercial tests generally meet standards of reliability,<sup>4</sup> which is necessary for comparability. However, commercial tests are not always aligned with the content that teachers are asked to teach, commercial tests may not include a pretest, and they may not have been designed for the purpose of teacher evaluation. Using a test which has not been validated for use in evaluating teachers could lead to inappropriate inferences about teacher effectiveness (Goe & Holdheide, 2011; Race to the Top Technical Assistance Network, 2011; Steele et al., 2010).<sup>5</sup> Alternatively, districts or states could develop their own assessments. Such assessments can be aligned with content standards, but effort is required to ensure that such assessments are reliable and valid for use in evaluating teachers (Goe & Holdheide, 2011; Steele et al., 2010).

## **Collective Measures of Student Growth**

In some districts and states, teachers in nontested subjects are evaluated based on aggregate measures of student growth, for example at the school or grade level. In Washington DC's IMPACT system, teachers in nontested subjects receive a school value-added score based on school growth in math and reading achievement on the state standardized assessments (District of Columbia Public Schools, 2011; Isenberg & Hock, 2011). This approach may encourage teachers to collaborate to improve student achievement (Steele et al., 2011; Race to the Top Technical Assistance Network, 2011), and it ensures comparability across schools (guideline 3). On the other hand, this approach does not provide a rigorous measure (or any measure) of students' knowledge in nontested subjects (guideline 3), and it does not differentiate among teachers (guideline 4). It masks the performance of both high- and low-performing teachers, which may lead to frustration if teachers feel they have limited ability to influence the evaluation

---

<sup>4</sup> A test is reliable if it is consistent over repeated measures.

<sup>5</sup> The validity of a test is judged relative to its purpose. A test that is valid for one purpose may not be valid for another. To assess the validity of a test used to evaluate teachers involves showing how well a teacher's score on the test represents the contribution of that teacher to student learning. For a more detailed discussion of validity in the context of teacher evaluation, see Steele et al., (2010).

measure and could even encourage teachers in nontested subjects to spend more time on subjects other than their own (Prince et al., 2009; Race to the Top Technical Assistance Network, 2011; Steele et al., 2011).

### **Multiple Approaches to Measuring Student Growth**

The three approaches to measuring student growth discussed above are not mutually exclusive. The Denver Public School district, for example, evaluates student growth for teachers in nontested grades using SLOs and also schoolwide achievement (Steele et al., 2008). As another example, Tennessee is currently using schoolwide student growth for teachers in nontested subjects, but plans to identify and develop alternative assessments for measuring growth for nontested subjects in the future (Tennessee Department of Education, 2011).

### **Other Measures Used to Assess Teacher Effectiveness**

Newly created teacher evaluation systems often use multiple measures to assess teacher effectiveness. The following sections, based largely on work by Goe, Bell, & Little (2008), summarize teacher effectiveness measures that are based primarily on teacher practice rather than student achievement. The advantages and disadvantages of each, with specific attention to guidelines 3 and 4, are discussed below.

#### **Classroom Observations**

Classroom observations are conducted by a trained rater who scores an instrument during a teacher observation. Instruments can differ in their developer (e.g., district-developed, developed by an outside vendor), the extent to which they are subject- or grade-specific, and the teaching domains covered. Classroom observation systems also differ in who can be a rater, the quality and quantity of training provided to raters, and how results are used (e.g., for professional development or tied to compensation). Two prominent classroom observation frameworks are Charlotte Danielson's *Enhancing Professional Practice: Framework for Teaching* and the University of Virginia's Classroom Assessment Scoring System (CLASS). Classroom observations can be rigorous and comparable (guideline 3) and differentiate among teachers (guideline 4) to the extent that they are based on a valid and reliable instrument that is designed to differentiate among teachers and that the instrument is scored by a well-trained rater. Classroom observations can provide detailed feedback potentially useful for evaluation (summative purposes) and also professional development (formative purposes). In addition, classroom observations can be conducted for teachers in a variety of subject areas, and one study finds a link between summary indices based on classroom observations and student achievement (Kane et al., 2011). On the other hand, effective training is essential for using instruments appropriately, and the evidence linking classroom observations to student outcomes is limited



(Goe et al., 2008). As an example at the state level, Rhode Island's teacher evaluation system includes, as one component, classroom observations.<sup>6</sup>

## **Portfolios**

A portfolio consists of materials that teachers select or create to show teaching practices and student performance. The teacher selects materials to fulfill criteria specified by the evaluator. Portfolios are often used in teacher preparation programs, and other states have adopted them as requirements for continuing or advanced licensing (e.g., Washington's ProTeach Portfolio program or Wisconsin's Master Educator License). The National Board for Professional Teaching Standards (NBPTS) offers teaching certificates in many areas, including eight in career and technical education, based on portfolio review.<sup>7</sup> Portfolios can provide detailed information on a teacher's practice, can show dimensions of teaching that are not observable in classroom observations (e.g., professionalism), and can be used for teachers in a wide variety of subjects (Goe et al., 2008). On the other hand, large-scale portfolio assessments have shown that raters scoring the same portfolio often assign different ratings,<sup>8</sup> and therefore comparability across classrooms could be difficult to achieve (guideline 3) (Goe et al., 2008). The extent to which portfolios are rigorous measures (guideline 3) that differentiate among teachers (guideline 4) depends on the design of the scoring rubric and, in the case of differentiation, the training of raters. In addition, the processes for compiling and evaluating portfolios are time-consuming (Goe et al., 2008).

## **Classroom Artifacts**

Instructional artifacts, which could include lesson plans, student work, tests, and scoring rubrics, can be judged for alignment with standards, comprehensiveness, clarity, and other criteria. Unlike portfolios, classroom artifacts are selected from materials that teachers use in their classrooms instead of from materials collected or created for the purpose of evaluation (Goe et al., 2008). There are a few structured protocols for analyzing classroom artifacts, including the Instructional Quality Assessment and the Scoop Notebook (Goe et al., 2008). While these two protocols cover math, reading, and science, other protocols could, in theory, be developed for other subjects. This type of measure can provide detailed information and may place a lower burden on teachers than compiling a portfolio. The extent to which classroom artifacts are rigorous and comparable measures (guideline 3) that differentiate among teachers (guideline 4) depends on the design of the scoring rubric and, for comparability and differentiation, the training of raters. Independent researchers have not validated protocols for analyzing classroom artifacts or linked scores to student outcomes (Goe et al., 2008).<sup>9</sup>

---

<sup>6</sup> See <http://www.ride.ri.gov/educatorquality/educatorevaluation/Docs/RIModelGuide.pdf> for more information.

<sup>7</sup> See [http://www.nbpts.org/userfiles/file/EAYA\\_CTE\\_Portfolio\\_Instructions\\_3162011.pdf](http://www.nbpts.org/userfiles/file/EAYA_CTE_Portfolio_Instructions_3162011.pdf) for portfolio instructions for the NBPTS certificate in CTE.

<sup>8</sup> More specifically, there are low rates of interrater reliability.

<sup>9</sup> Classroom artifacts and portfolios could, theoretically, be used to demonstrate student growth over time (Goe & Holdheide, 2011; Reform Support Network, 2011). However, we could not identify districts or states that currently

## Student Surveys

As discussed by Goe and colleagues (2008), student surveys could play a role in teacher evaluation. Evidence on the reliability of student ratings and their relationship with student achievement is limited, but students are a key stakeholder and may be able to provide information not easily measured by other types of assessments (Goe et al., 2008). While research evidence is limited, a small number of studies suggest that secondary students may provide reliable ratings of teachers and that their ratings may link to student achievement (see discussion in Goe et al., 2008). On the other hand, at the postsecondary level, where student ratings are often part of tenure decisions, the evidence for the usefulness of student surveys is mixed. Some evidence suggests that professors may hold students to less rigorous standards in order to boost ratings (Wilson, 1998). Moreover, students are not typically qualified to evaluate teachers on some aspects of teaching, including content knowledge (Goe et al., 2008). Comparability (guideline 3) would be difficult to achieve if students rate teachers relative to other teachers in the same school, and at least one study found that student responses skewed toward higher satisfaction (Peterson, Wahlquist, & Bone, 2000), making it more difficult to distinguish among levels of teacher effectiveness (guideline 4).

## Summary

Consistent with other district, state, and federal policy initiatives, Michigan's teacher evaluation system requires that teacher evaluations use multiple measures, that the measures include a student growth component, that the measures are rigorous and comparable, and that the measures differentiate teachers into levels of effectiveness. This overview has discussed general approaches to evaluating student growth for teachers in nontested subjects and also has provided information on additional types of measures that have been used in teacher evaluations.

---

use the measures in that way, and such a use would be susceptible to the disadvantages already discussed, as well as potentially new ones.

**Table 1. Summary of State Policies on the Evaluation of Teachers in Nontested Subjects**

State	Evaluation of CTE Educators	Measures of Student Growth in Nontested Subjects
<b>Illinois</b>	State evaluation policies do not outline specific requirements or measures for CTE educators.	The Performance Evaluation and Reform Act (PERA) assigned the Illinois State Board of Education (ISBE) the responsibility of defining minimum criteria for the evaluation systems that districts will be required to adopt under the new law. Districts will be able to choose the specific measures that will proxy for student growth and targets for student growth, including for different student subgroups, as well as the percentage of effectiveness ratings that will be determined by measures of student growth, as long as the measures and percentage meet or exceed the minimum criteria established by ISBE. ISBE tasked the Performance Evaluation Advisory Committee (PEAC) with creating or determining methods to measure student growth in all subjects and grades, not only those that require standardized testing. In a recent presentation posted on its website, PEAC's draft criteria for the evaluation systems require districts to use at least one measure of student growth based on an assessment that is "rigorous, aligned with the course's curriculum, and that the evaluator and teacher determine measures student learning." Measures of growth should demonstrate the "change in a student's learning between two or more points in time" and must comprise at least 30% of the final rating. The examples provided are teacher-created assessments and assessments of student performance. (Slide 13, <a href="http://www.isbe.state.il.us/peac/pdf/ISBE-PEAC_overview_pres102011.pdf">http://www.isbe.state.il.us/peac/pdf/ISBE-PEAC_overview_pres102011.pdf</a> )
<b>Indiana</b>	State evaluation policies do not outline specific requirements or measures for CTE educators.	Senate Bill 1 requires local school corporations (LSC) to implement new teacher evaluation systems based on multiple measures of teacher effectiveness, including student growth measures. IDOE will define minimum standards for designing teacher evaluation systems at the LSC level. LSCs must include objective and rigorous measures of student achievement for nontested subjects. IDOE is in the process of developing a model teacher effectiveness rubric called RISE to evaluate teachers annually on all criteria established in Senate Bill 1, except for student growth measures.
<b>Iowa</b>	State evaluation policies do not outline specific requirements or measures for CTE educators.	Iowa's current teacher evaluation system is based on eight teaching standards for different stages of a teacher's career. In advance of submitting its Race to the Top application, Iowa amended one of the eight teaching standards on which teachers are evaluated to include a requirement that teachers provide an analysis of student learning and growth based on teacher-created tests and authentic measures, as well as any standardized and districtwide tests (281 Iowa Administrative Code 83.4(7)(a)(5)).
<b>Michigan</b>	State evaluation policies do not outline specific requirements or measures for CTE educators.	Public Act No. 205 requires districts to design and implement a "rigorous, transparent, and fair performance evaluation system" for teachers that includes measures of student growth as a significant factor. School districts must develop a clear approach for measuring student growth using national, state, or local assessments or other "objective criteria." Public Act No. 102 requires that a Governor's Council on Educator Effectiveness convene by October 31, 2011. The Council must produce a report by April 30, 2012, with recommendations for a model state teacher and administrator evaluation tool that includes a student growth/assessment tool. Districts will implement that tool or a tool that is consistent with the state-provided model beginning in the 2013–14 school year.

State	Evaluation of CTE Educators	Measures of Student Growth in Nontested Subjects
<b>Minnesota</b>	State evaluation policies do not outline specific requirements or measures for CTE educators.	In July 2011, Minnesota passed House File No. 26, which revised the statewide teacher evaluation system in Minnesota. The legislation allows local school districts to design their own evaluation systems that meet specific state-mandated criteria or implement a model evaluation system that the Minnesota Department of Education (MDOE) will develop. House File No. 26 stipulates that 35 percent of the teacher effectiveness rating be based on teacher value-added assessment models and/or state and local measures of student growth.
<b>New York</b>	State evaluation policies do not outline specific requirements or measures for CTE educators.	Measures of student growth on a state assessment or a comparable measure of student growth constitutes 20 percent of the evaluation of an educator whose subject or grade does not have a value-added model. Districts should use a state assessment for this measure, if one exists or a Regent equivalent. If one does not exist, the District can choose an assessment from list of state-approved third party assessments and Regents equivalents. Alternatively, the district may develop assessments, as long as the district verifies comparability and rigor; use school-wide, group, or team results based on state assessments; or school or teacher-created assessment (for more information, please see: <a href="http://usny.nysed.gov/rttt/teachers-leaders/greenmemo.pdf">http://usny.nysed.gov/rttt/teachers-leaders/greenmemo.pdf</a> ). Locally selected measures of student learning will constitute 20 percent of an educator's evaluation in subjects or grades for which there are no value-added models. Districts may select the measures they would like to use, but the measures must be rigorous and comparable across classrooms in the same grade/subject. The state has provided a list of approved assessments, which are aligned with the current New York State Learning Standards and have been evaluated based on criteria established in the state's Request for Qualification. For a complete list of approved assessments, as well as additional information on assessments of student growth, please see: <a href="http://usny.nysed.gov/rttt/teachers-leaders/assessments/#assess">http://usny.nysed.gov/rttt/teachers-leaders/assessments/#assess</a>
<b>Ohio</b>	State evaluation policies do not outline specific requirements or measures for CTE educators.	In July 2009, the former governor of Ohio signed House Bill 1, also known as the Education Reform Plan, into law. The bill directed the Educator Standards Board to develop a method for measuring student academic growth over a one-year period and incorporating those measures into a recommended model Ohio Teacher Evaluation System (OTES). The OTES model is intended to act as a guide for districts in the design and implementation of a high-quality, performance-based teacher evaluation system. The most recent draft of the OTES model utilizes multiple measures of teacher effectiveness, including student growth, as 50 percent of the evaluation rating. It also states that an ODE cross-agency team is still in the process of developing measures of student academic growth for grade levels and subjects for which the value-added progress dimension does not apply (Ohio Teacher Evaluation System, p. 20) Please see: <a href="http://education.ohio.gov/GD/Templates/Pages/ODE/ODEDetail.aspx?page=3&amp;TopicRelationID=521&amp;ContentID=108217">http://education.ohio.gov/GD/Templates/Pages/ODE/ODEDetail.aspx?page=3&amp;TopicRelationID=521&amp;ContentID=108217</a> ).

State	Evaluation of CTE Educators	Measures of Student Growth in Nontested Subjects
Tennessee	<p>35 percent of a CTE teacher's effectiveness rating in the new evaluation model created through the First to the Top Act of 2010 will be determined by student growth. Until the Tennessee Department of Education finishes developing comparable, alternative growth measures for students in nontested subjects and grades, this portion of the CTE teacher's effectiveness rating will be based on an overall school-level growth measure, or a school-level growth measure in literacy, math, or math/literacy. In addition, 15% of a teacher's effectiveness rating will be determined by measures of student achievement. CTE teachers will consult with their administrators to choose a measure of student achievement from an Approved Measures Matrix created by the State Board of Education that is closely aligned with their primary responsibilities. The categories of approved measures for CTE teachers are discipline-specific state assessments, off-the-shelf tests, post-secondary matriculation, persistence, or placement, participation in advanced coursework, or the ninth-grade promotion and retention rate. For more information, please see: <a href="http://team-tn.org/teacher-model">http://team-tn.org/teacher-model</a></p>	<p>Under Tennessee's new evaluation model created through the First to the Top Act of 2010, 35 percent of an educator's overall evaluation must be measures of student growth. Approximately 45 percent of Tennessee's educators have an individual teacher effect score from the Tennessee Value-Added Assessment System (TVAAS). In order to meet the requirement that the evaluation system be implemented in fall 2011, most teachers in nontested subjects and grade levels will use a school-level growth score for the 2011-12 school year (either overall, literacy, math, or math/literacy). The Tennessee Department of Education is still in the process of identifying, developing, and approving comparable, alternative growth measures in nontested subjects and grades. Though pilots are being conducted for most educator groups, alternate growth measures under pilot will not be available for inclusion in the evaluation model until the 2012-13 school year, pending their approval. Most school buildings have TVAAS schoolwide scores, but Pre-K–2 and alternative schools do not have this type of TVAAS information. For this coming year, the school-level value-added composites for Pre-K–2 schools will come from the feeder schools; alternative schools will use districtwide value-added in lieu of school-level data. In addition, 15% of a teacher's effectiveness rating will be determined by measures of student achievement. The State Board of Education has provided an Approved Measures Matrix, from which teachers and administrators can choose a measure that is aligned as closely as possible to the particular teacher's primary responsibility. The categories of approved measures are discipline-specific state assessments, off-the-shelf tests, post-secondary matriculation, persistence, or placement, participation in advanced coursework, or the 9th grade promotion and retention rate. For more information, please see the following links:</p> <ul style="list-style-type: none"> <li>• Approved Achievement Measures Matrix: <a href="http://team-tn.org/assets/educator-resources/Approved_Achievement_Measures_Matrix.pdf">http://team-tn.org/assets/educator-resources/Approved_Achievement_Measures_Matrix.pdf</a></li> <li>• Approved Achievement Measures: Additional Information: <a href="http://team-tn.org/assets/educator-resources/Approved_Achievement_Measures_Expanded_Options.pdf">http://team-tn.org/assets/educator-resources/Approved_Achievement_Measures_Expanded_Options.pdf</a></li> <li>• Achievement Measure Worksheet: <a href="http://team-tn.org/assets/educator-resources/Achievement_Measure_Worksheet.pdf">http://team-tn.org/assets/educator-resources/Achievement_Measure_Worksheet.pdf</a></li> <li>• Example Achievement Measure Worksheet Visit: <a href="http://team-tn.org/assets/educator-resources/Example_Achievement_Measure_Worksheet.pdf">http://team-tn.org/assets/educator-resources/Example_Achievement_Measure_Worksheet.pdf</a></li> <li>• Sample Calculations: <a href="http://team-tn.org/assets/educator-resources/Sample_Calculation_for_Achievement_Measures.pdf">http://team-tn.org/assets/educator-resources/Sample_Calculation_for_Achievement_Measures.pdf</a></li> <li>• Growth for Non-Tested Subjects and Grades Summary: <a href="http://team-tn.org/assets/educator-resources/Non-Tested_Subjects_and_Grades_Summary.pdf">http://team-tn.org/assets/educator-resources/Non-Tested_Subjects_and_Grades_Summary.pdf</a></li> <li>• Non-tested Subjects and Grades Q&amp;A: <a href="http://team-tn.org/assets/educator-resources/Non-tested_subjects_and_grades_QandA.pdf">http://team-tn.org/assets/educator-resources/Non-tested_subjects_and_grades_QandA.pdf</a></li> <li>• Measuring Student Learning: <a href="http://team-tn.org/assets/educator-resources/TN_Evaluation_Measuring_StudentLearning_September_2011.pdf">http://team-tn.org/assets/educator-resources/TN_Evaluation_Measuring_StudentLearning_September_2011.pdf</a></li> </ul>

State	Evaluation of CTE Educators	Measures of Student Growth in Nontested Subjects
Wisconsin	<p>State evaluation policies do not outline specific requirements or measures for CTE educators.</p>	<p>Wisconsin has not yet developed measures of student growth for nontested subjects. Wisconsin’s State Superintendent convened the Educator Effectiveness Design Team to make recommendations for reforming Wisconsin’s educator effectiveness system. The Design Team released a preliminary report in November 2011 (<a href="http://dpi.wi.gov/tepd/pdf/ee_report_prelim.pdf">http://dpi.wi.gov/tepd/pdf/ee_report_prelim.pdf</a>). The report contains recommendations for changes to state policies, but no policy changes have been made to date.</p> <p>The Design Team recommended that multiple measures of student outcomes be used to evaluate teachers:</p> <ol style="list-style-type: none"> <li>1. Individual value-added data on statewide standardized assessments (currently grades 3-7 reading and math);</li> <li>2. District-adopted standardized assessment results</li> <li>3. Student Learning Objectives (SLOs) established by teachers with approval from administrators. SLOs are a participatory method of setting measurable goals around measurable growth in student performance during the course of instruction. SLOs can be based on teacher-developed or other classroom assessments.</li> <li>4. District choice of data based on improvement strategies and aligned to school and district goals based on areas of need highlighted by the state accountability system.</li> <li>5. For elementary and middle school levels, school-wide reading scores will be used.</li> </ol> <p>The Design Team recommended that all five types be used when available and specified the way the measures would be weighted. For teachers of grades and subjects that are not assessed through standardized assessments, SLOs will be the primary means of measuring student growth.</p>

## References

- Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25(1), 95–135.
- Brandt, C., Mathers, C., Oliva, M., Brown-Sims, M., & Hess, J. (2007). *Examining district guidance to schools on teacher evaluation policies in the Midwest region* (Issues & Answers Report, REL 2007-No. 030). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, REL Midwest.
- Braun, Henry I. *Using student progress to evaluate teachers: A primer on value-added models*. Princeton, NJ: Educational Testing Service.
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, 8(1): 1-44.
- Denver Public Schools. (n.d.). Student growth objective and monitoring process: Elementary/ECE-Grade 8/ Secondary. Retrieved October 15, 2011, from <http://static.dpsk12.org/gems/sgo/SGOMASTERTeacher200809.pdf>
- District of Columbia Public Schools. (2011 ). Value-Added: Overview. Retrieved November 2, 2011, from [http://dcps.dc.gov/DCPS/In+the+Classroom/Ensuring+Teacher+Success/IMPACT+\(Performance+Assessment\)/Value-Added](http://dcps.dc.gov/DCPS/In+the+Classroom/Ensuring+Teacher+Success/IMPACT+(Performance+Assessment)/Value-Added)
- Glazerman, S., Goldhaber, D., Loeb, S., Raudenbush, S., Staiger, D.O., & Whitehurst, G.J. (2011). *Passing Muster: Evaluating Teacher Evaluation Systems*. Brown Center on Education Policy at Brookings.
- Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved October 15, 2011, from <http://www.tqsource.org/publications/EvaluatingTeachEffectiveness.pdf>
- Goe, L., & Holdheide, L. (2011). *Measuring teachers' contributions to student learning growth for nontested grades and subjects*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved October 15, 2011, from <http://www.tqsource.org/publications/MeasuringTeachersContributions.pdf>
- Hanushek, E.A., & Rivkin, S.G. Generalizations about using value-added measures of teacher quality. *American Economic Review*, 100(2): 267-71.
- Holdheide, L.R., Goe, L., Croft, A., & Reschly, D.J. (2010). *Challenges in evaluating special education teachers and English language learner specialists*. Washington, DC: National



Comprehensive Center for Teacher Quality. Retrieved November 10, 2011, from <http://www.tqsource.org/publications/July2010Brief.pdf>

Isenberg, E. & Hock, H. (2011). Design of Value-Added Models for IMPACT and TEAM in DC Public Schools, 2010–2011 School Year. Washington, DC: Mathematica Policy Research. Retrieved November 2, 2011, from

<http://dcps.dc.gov/DCPS/Files/downloads/In-the-Classroom/Design%20of%20Value-Added%20Models%20for%20DCPS%202010-2011-FINAL.pdf>

Jacob, B.A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26(1), 101-136.

Kane, T.J., Taylor, E.S., Tyler, J.H., & Wooten, A.L. Evaluating teacher effectiveness: Can classroom observations identify practices that raise achievement? *Education Next*, Summer: 55-60.

Mathers, C., Oliva, M., & Laine, S.W.M. (2008). Improving instruction through effective teacher evaluation: Options for states and districts. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved November 10, 2011, from

<http://www.tqsource.org/publications/February2008Brief.pdf>

Metz, E. & Wraight, S. (2011). Summary of current teacher evaluation policies and initiatives in Midwest region states. Naperville, IL: REL Midwest.

The New Teacher Project (2010). Teacher Evaluation 2.0. New York: The New Teacher Project.

Nye, B., Konstantopoulos, S., & Hedges, L.V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, 26(3): 237-257.

Peterson, K.D., Wahlquist, C., & Bone, K. Student surveys for school teacher evaluation. *Journal of Personnel Evaluation in Education*, 14(2):135-153.

Potemski, A., Baral, M., & Meyer, C. 2011. Alternative measures of teacher performance. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved October 15, 2011, from

[http://www.tqsource.org/pdfs/TQ\\_Policy-to-PracticeBriefAlternativeMeasures.pdf](http://www.tqsource.org/pdfs/TQ_Policy-to-PracticeBriefAlternativeMeasures.pdf)

Prince, C. D., Schuermann, P. J., Guthrie, J. W., Witham, P. J., Milanowski, A. T., & Thorn, C. A. (2006). The other 69 percent: Fairly rewarding the performance of teachers of nontested subjects and grades. Washington, DC: U.S. Department of Education, Office of Elementary and Secondary Education.

Prince, C.D., Schuermann, P.J., Guthrie, J.W., Witham, P.J., Milanowski, A.T., Thorn, C.A. (2009). The other 69 percent: Fairly rewarding the performance of teachers of nontested



subjects and grades. Revised edition. Center for Education Compensation Reform. Retrieved October 15, 2011, from <http://www.cecr.ed.gov/pdfs/guide/other69Percent.pdf>

Reform Support Network. (2011). Great teachers and leaders: State considerations on building systems of educator effectiveness. Fairfax, VA: ICF International. Retrieved October 15, 2011, from <http://www.ksde.org/Portals/0/Growth%20ModelsN/Teacher%20Evaluations.pdf>

Rockoff, J.E. (2004). The impact of individual teachers on student achievement. Evidence from panel data. *American Economic Review*, 94(2), 247-252.

Race to the Top Technical Assistance Network. (2011). Measuring student growth for teachers in non-tested grades and subjects: A primer. Fairfax, VA: ICF International. Retrieved October 15, 2011, from [http://www.swcompcenter.org/educator\\_effectiveness2/NTS\\_PRIMER\\_FINAL.pdf](http://www.swcompcenter.org/educator_effectiveness2/NTS_PRIMER_FINAL.pdf)

Rhode Island Department of Education. (2011). Guide to Measures of Student Learning for Administrators and Teachers 2011-2012. Retrieved November 2, 2011, from <http://www.ride.ri.gov/educatorquality/educatorevaluation/Docs/GuideSLO.pdf>

Steele, J.L., Hamilton, L.S., & Stecher, B.M. (2010). Incorporating student performance measures into teacher evaluation systems. Santa Monica, CA: RAND Corporation. Retrieved November 1, 2011, from [http://www.rand.org/content/dam/rand/pubs/technical\\_reports/2010/RAND\\_TR917.pdf](http://www.rand.org/content/dam/rand/pubs/technical_reports/2010/RAND_TR917.pdf)

Toch, T. & Rothman, R. (2008). Rush to judgment: Teacher evaluation in public education. Washington, DC: Education Sector. Retrieved November 1, 2011, from [http://www.educationsector.org/sites/default/files/publications/RushToJudgment\\_ES\\_Jan08.pdf](http://www.educationsector.org/sites/default/files/publications/RushToJudgment_ES_Jan08.pdf)

TAP. (n.d.). Performance Based Compensation. Retrieved November 2, 2011, from <http://www.tapsystem.org/action/action.taf?page=pbcc>

Tennessee Department of Education. (2011). Measuring Student Growth in Non-tested Grades and Subjects. Retrieved November 2, 2011 from [http://team-tn.org/assets/educator-resources/Non-Tested\\_Subjects\\_and\\_Grades\\_Summary.pdf](http://team-tn.org/assets/educator-resources/Non-Tested_Subjects_and_Grades_Summary.pdf)

Wilson, R. (1998, January 16). New research casts doubt on value of student evaluations of professors. *The Chronicle of Higher Education*, p. A12.